

Scene Change Aware Intra-Frame Rate Control for H.264/AVC

Wen-Jiin Tsai, *Member, IEEE*, and Ting-Li Chou

Abstract—Most of rate-control research focuses on inter-coded frames, instead of intra-coded frames which are more possible to cause the problem of buffer overflow. This letter presents a rate control algorithm for intra-frame coding. We propose a Taylor-series-based rate-QS model and a scene-change aware rate-QS model to determine quantization parameters for general intra-frames and scene-change frames, respectively. Simulation results show that compared to competed approaches, the proposed method achieves better and stable quality with low buffer fullness.

Index Terms—H.264, intra frames, prediction model, rate control.

I. INTRODUCTION

Intra-only coding schemes for professional applications have been standardized as part of H.264/AVC profiles [1]. For example, in the 7th edition of H.264/AVC specification, there are three profiles: *High10Intra*, *High 4:2:2 Intra*, and *High 4:4:4 Intra*, that are intra-only related. Since intra-only coding schemes do not exploit temporal correlation between frames, they have the advantages with regard to convenient editing, parallel processing, and so on, over group-of-picture (GOP) based coding schemes. Besides, intra-only coding is error robust in comparison to GOP-based coding schemes; even though GOP-based schemes are added with error-resilient coding techniques such as layered coding [2], multiple description coding [3], and so on. Because of these features, it is greatly appropriate to adopt intra-only compression for the high-end applications.

Rate control techniques have been studied intensively for many coding standards. The challenge of rate control in video encoding is to determine appropriate quantization parameters (QP) to achieve the best video quality within the given bit-rate constraint. Li *et al.* [4] proposed an efficient rate control algorithm, which has been adopted by JVT in the latest H.264/AVC reference software [5] and known as the JVT-G012 [4]. In JVT-G012, the QPs for P-frames are determined by the rate-quantization quadratic model [6]; while the QP for an I-frame is decided by the average QP of all P-frames in

the previous GOP. The starting QP of the I-frame in the first GOP depends on the value of bit-per-pixels. However, without taking buffer status and frame complexity into considerations, this approach usually allocates too much bits for I-frames and thus, degrades the video quality of succeeding P-frames due to insufficient bit-rates. Plenty of rate control algorithms have been proposed to improve the performance of JVT-G012. However, most of them focused on inter-coding instead of intra-coding even though intra-coded frames are more likely to cause buffer overflow.

The quadratic model used in JVT-G012 assumes that discrete cosine transform (DCT) coefficients are a Laplacian distribution, which is widely used to model the relation between bit-rates and QPs for inter-coded frames [7]. In [8], Kamaci *et al.* proposed a solution which uses a Cauchy probability density function for DCT coefficient estimation. They have shown that the Cauchy model outperforms Laplacian model for both intra-frames and inter-frames. Kamaci *et al.* further presented a Cauchy-based intra-frame rate estimation model which is utilized to approximate the entropy function of quantization. However, their approach adopts a simple scheme for model parameter updating and therefore, is not sufficiently adaptive to the change of frame complexity. Based on Kamaci *et al.*'s rate estimation model, Jing *et al.* [9] proposed an improved model which has sufficient adaptability to the varying complexity of intra frames. They revised Kamaci *et al.*'s rate estimation model by adding the complexity measure of an intra frame, which is defined as the average gradient-per-pixel of that frame.

This letter presents a new scheme to determine QPs for general intra-frames (I_G frame) and scene-change intra-frames (I_{SC} frame). Instead of using the way adopted in JVT-G012, we proposed a rate control algorithm taking account for frame complexity to decide proper QPs for both types of frames. Experimental results show that the proposed approach performs better than competed methods. The remainder of this letter is organized as follows. Section II presents the proposed rate control scheme and Section III provides the simulation results. Finally, Section IV concludes this letter.

II. PROPOSED INTRA-FRAME RATE-CONTROL

In this section, we first propose a Taylor-series-based rate-QS model for I_G frames, and then a scene-change aware rate-QS model for I_{SC} frames. The proposed Taylor-series-based

Manuscript received August 10, 2009; revised February 12, 2010 and April 28, 2010; accepted August 25, 2010. Date of publication October 14, 2010; date of current version January 22, 2011. This paper was recommended by Associate Editor W. Gao.

W.-J. Tsai is with the Department of Computer Science, National Chiao Tung University, Hsinchu 300, Taiwan (e-mail: wjtsai@cs.nctu.edu.tw).

T.-L. Chou is with the Department of Core Technology, Cyberlink Corporation, Taipei 231, Taiwan (e-mail: david74.cs96g@g2.nctu.edu.tw).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TCSVT.2010.2087473

model is based on Jing's rate estimation model [9] as follows:

$$R = G \times a \cdot QS^b \quad (1)$$

where QS is quantization step size, b is a constant, a is updated frame by frame, and G is calculated as

$$G = \frac{1}{M \cdot N} \left(\sum_{i=0}^{M-1} \sum_{j=0}^{N-1} |I_{i,j} - I_{i+1,j}| + |I_{i,j} - I_{i,j+1}| \right) \quad (2)$$

where M and N are the horizontal and vertical sizes of the frame, respectively. $I_{i,j}$ is the luminance value of the pixel at location (i, j) . Jing's method of deciding parameter a assumes that a is stationary between frames, so it predicts current a by using the a from previously encoded frames. However, parameter a is not always stationary. To solve this problem, we rewrite (1) and define $R_{norm,i}$, the *normalized bitrate* of the i th frame, as

$$R_{norm,i} = \frac{R_i}{G} = aQS^b. \quad (3)$$

To explore the relationship between R_{norm} and QS in this equation, we gathered the statistics of R_{norm} from different frames intra-encoded with all possible QS s. Fig. 1 shows the R_{norm} as a function of QS for the first five frames in *Akiyo* sequence. It is observed that the $R_{norm}(QS)$ curves of consecutive frames are almost identical. In other words, we can assume $R_{norm,i}(QS) = R_{norm,i-1}(QS)$. Although it is not likely to have the whole $R_{norm}(QS)$ curve of the previous frame, there is one single point available because the QS and R used in the previous frame can be obtained after it has been encoded. Based on Taylor-series theory [10] that any infinitely differentiable function, $f(x)$, can be represented as an infinite sum of terms calculated from all the values of derivation at a single point, we rewrite (3) for a given QS as

$$\begin{aligned} R_i(QS) &= G_i \times R_{norm,i}(QS) = G_i \times R_{norm,i-1}(QS) \\ &= G_i \times \left[R_{norm,i-1} + \frac{b \left(\frac{R_{norm,i-1}}{QS_{i-1}} \right)}{1!} (QS - QS_{i-1}) \right. \\ &\quad \left. + \frac{b \cdot (b-1) \left(\frac{R_{norm,i-1}}{QS_{i-1}^2} \right)}{2!} (QS - QS_{i-1})^2 \right] \quad (4) \end{aligned}$$

where b is a constant the same as that in (3). To obtain this b , we analyzed over 3000 frames from various sequences by intra-encoding each frame using all possible QS s. By plotting the encoded bitrates, R , as a function of QS , we found that these experimental points can be fit in with a curve using a constant exponent -0.76 . Thus, we set $b = -0.76$ in (3) and (4).

In (4), since $R_{norm,i-1}$ and QS_{i-1} are available when encoding frame i , the $R_i(QS)$ for any given QS can be obtained before frame i is encoded. It is reliable because unpredictable a no longer exists. With this model, the best QP for an I_G frame, subject to its bit budget, can be determined as follows. Take every candidate QS as an input, calculate the predicted bit-rate of this frame using (4), choose the QS associated with the predicted bit-rate closest to its bit budget and finally, the

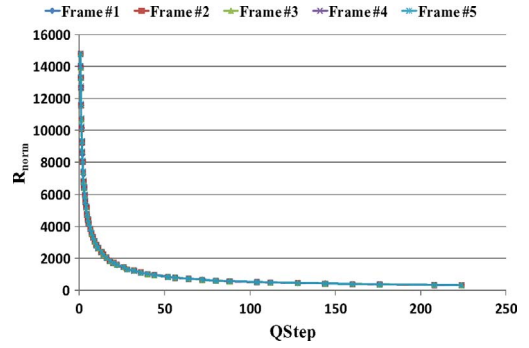


Fig. 1. Curves between R_{norm} and QS of *Akiyo* frames.

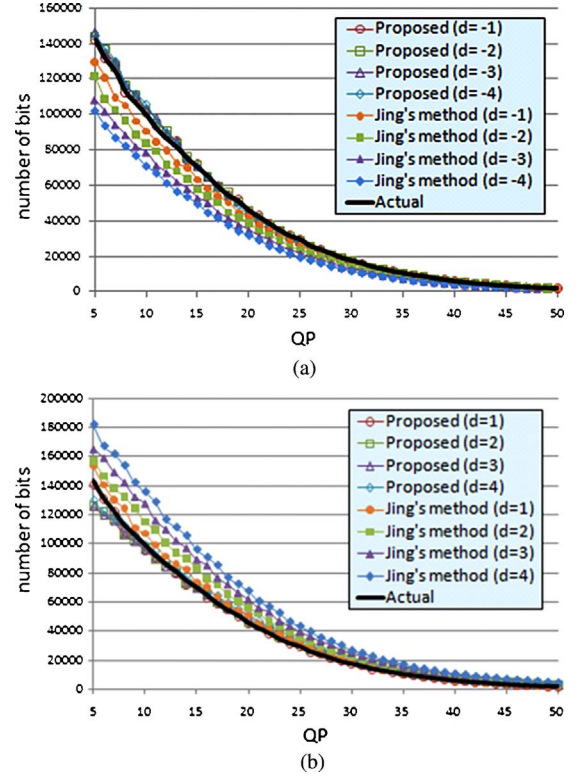


Fig. 2. Prediction result for the fifth frame of *Foreman* sequence. (a) QP difference, d , is negative. (b) QP difference, d , is positive.

QP corresponding to the QS can be decided. Since all frames are of the same type in intra-only compression, a simple and efficient bit budget allocation is to give each frame the same bit-rate, that is

$$R_t = \frac{R_{remain}}{N_r} \quad (5)$$

where R_{remain} is the available bit rate and N_r is the number of remaining frames. To prevent large peak signal-to-noise ratio (PSNR) deviation, we limit the maximum QP difference between successive frames to 4.

To show the correctness of the proposed model, Fig. 2 gives an example of prediction results for the fifth frame of *Foreman* sequence, where the d denotes the QP difference between the fifth frame and its previous frame. It shows that the proposed model achieves a more accurate bit-rate estimate than Jing's method which suffers from large prediction error as the absolute value of d increases. The result stems from

the fact that Jing's method depends on the parameter a which is predicted from previous encoded frame and hence, it will predict less bits when d is negative and more bits when d is positive. The proposed model did not have such a problem because it depends on R_{norm} curve similarity between successive frames.

When scene change happens on a frame, the information from previous coded frames is inadequate to predict the coding result of this frame. Here, a *scene-change aware rate-QS model* is proposed to predict the bit-rate for such scene-change frames (I_{SC} frames). We have analyzed over 3000 frames from various sequences by intra-encoding each frame using all possible QSs and calculated the parameter a of (1) for each frame using the encoded bitrates, R , the measured G , and the constant $b = -0.76$. By multiplying the calculated a by G for each frame and plotting them as a function of G on a figure, we found that these experimental points (i.e., the relation between aG and G) can be fit in with a straight line. According to this result, we re-write (1) as

$$R = (\omega \cdot G + \mu)QS^b \quad (6)$$

where ω and μ are constants, 6022.1 and 885 220, respectively, for QCIF sequences. For CIF sequences, they are $\omega = 27\,360$, $\mu = 338\,726$; and for SD sequences, they are $\omega = 7702.9$, $\mu = 2E + 06$. The b is set to -0.76 as that in (1), regardless of frame resolutions. With this model, since ω , μ , and b are all constants and the frame complexity G can be obtained from current frame itself, the QS subject to the target bit-rate R_t can be calculated and then the corresponding QP can be determined. Namely, by using (6), the QP of an I_{SC} frame can be decided by all the information from that frame itself, nothing from previous frames is required.

To examine the effects of using scene-change aware model in rate control, Fig. 3 shows the prediction error for intra-only compression on a sequence cascaded by *Hall* and *Foreman* sequences with scene changes at frames 10, 20, 30, and 40. The bit-rate is set to 1024 kb/s. The methods used for comparison include: Jing's method, the proposed Taylor-series-based model (T model), and the T model integrated with the scene-change aware model (T+SC). In Fig. 3, it is obvious to see that, compared with Jing's method and T model, the T+SC method is more accurate on bit-rate prediction at scene-change frames.

The intra-rate control algorithm with the proposed two models is summarized as follows: after the input of an intra frame, calculate the complexity G and the target bit-budget R_t of this frame first, according to (2) and (5), respectively. Then, perform scene-change detection to see whether this frame is an I_{SC} frame or not. If it is an I_{SC} frame, adopt scene-change aware model to determine its QP; otherwise, adopt Taylor-series-based model. With the decided QP, perform H.264/AVC RDO and the ensuing encoding process for this frame. Finally, update parameter $R_{\text{norm},i-1}$ if the end of the sequence is not reached. Note that since our two models are independent of scene-change detection methods, any algorithm that can correctly detect scene changes can be incorporated into our approach.

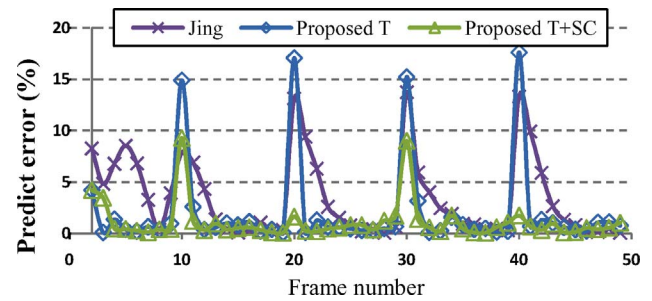


Fig. 3. Bit-rate prediction comparison on a cascading sequence.

III. EXPERIMENTAL RESULTS

Our rate control algorithm has been integrated into JVT reference software JM15.0 [5]. The simulation was produced with three QCIF (*Foreman*, *Mobile*, and *News*), three CIF (*Foreman*, *Mobile*, and *News*), and three 704×576 SD (*Crew*, *Ice*, and *Soccer*) sequences. These sequences differ from those used in the training stage for calculating parameters, b , ω , and μ . In addition, to test the proposed method under scene change conditions, four sequences, *Combo1* (QCIF *Trevor-Stephan-Silent-Coastguard*), *Combo2* (QCIF *Akiyo-Mobile*), *Combo3* (CIF *Foreman-Mobile-News*), and *Combo4* (SD *Crew-Ice-Soccer*) were created by cascading corresponding sequences, and the intervals of every two consecutive scene cuts are 50 frames long. As for scene-change detection, the method based on the difference of histogram [11] is adopted. To see the effects of the proposed two models separately, we use "T" to denote the rate control method with proposed Taylor-series-based model only, and "T+SC" to denote the T approach together with the scene-change aware model. We compare the proposed methods with Jing's method [9] and the intra-only rate control algorithm in JM15.0, where the later one is a modified version based on G012 [4]. The overall performance results are measured in terms of actual bit-rate, average PSNR, and PSNR standard deviation (*StdDev*). The PSNR *StdDev* of a video sequence is defined by

$$StdDev = \sqrt{\frac{1}{K} \sum_{i=1}^K (x_i - \bar{x})^2} \quad (7)$$

where x_i is the PSNR measured from each frame, K is the number of frames, and \bar{x} is the mean of PSNRs, calculated as

$$\bar{x} = \frac{1}{k} \sum_{i=1}^k x_i.$$

The experimental results are summarized in Table I. It is observed that the performance gain of the proposed T method over Jing's method is small. Although we have shown in Fig. 2 that our Taylor-series-based model can achieve a much more accurate bit-rate prediction than Jing's model, the performance gain from this seems quite limited. This is due to that although Jing's method has inaccurate bit-rate prediction for the QP of a frame when this QP differs largely from the QP of its previous frame, the QP differences between successive frames are within 2 in most cases. Thus, the prediction error of Jing's method is tolerable and able to select proper QPs.

TABLE I
PERFORMANCE COMPARISONS FOR INTRA-ONLY COMPRESSION

Target Bit-rate	Seq.	Bit-rate				Avg. PSNR (dB)				PSNR StdDev			
		JM	Jing	T	T+SC	JM	Jing	T	T+SC	JM	Jing	T	T+SC
512 kbps (QCIF)	Foreman	512.02	512.01	512.44	511.92	34.61	34.63	34.64	35.39	3.719	3.714	3.704	0.462
	Mobile	512.44	512.37	512.91	511.96	21.53	21.53	21.52	23.73	6.944	7.09	6.884	0.652
	News	510.48	512.15	512.09	512.14	32.7	32.75	32.75	33.76	4.46	4.535	4.367	0.502
	Combo1	511.9	512.2	512.01	512.04	31.33	31.41	31.40	31.78	4.473	4.322	4.314	3.236
	Combo2	511.65	511.9	512.11	512.22	30.76	30.88	30.90	31.44	8.154	8.113	8.078	7.634
	Average	511.698	512.126	512.312	512.056	30.186	30.24	30.242	31.22	5.55	5.5548	5.4694	2.4972
2048 kbps (CIF)	Foreman	2046.44	2048.7	2048.04	2048.7	38.09	38.28	38.28	38.28	1.009	0.942	0.943	0.908
	Mobile	2046.92	2047.6	2048.01	2048.1	24.78	24.96	24.96	25.49	3.843	3.825	3.706	0.627
	News	2047.8	2049.5	2048.29	2048.3	39.07	39.3	39.31	39.35	1.036	0.935	0.93	0.387
	Combo3	2048.46	2047.8	2047.3	2048.9	33.75	34.14	34.16	34.17	5.94	5.534	5.462	5.419
	Average	2047.4	2048	2047.9	2048	33.92	34.17	34.18	34.323	2.957	2.809	2.76	1.835
4096 kbps (SD)	Crew	4109.73	4094.1	4100.39	4100.6	37.8	38.23	38.25	38.3	0.989	1.059	0.999	0.832
	Ice	4083	4096.6	4097.62	4098.1	42.76	42.86	42.88	42.93	1.183	1.225	1.143	0.817
	Soccer	4034.93	4096.6	4093.49	4095.5	33.31	33.86	33.86	33.83	2.282	2.235	2.237	2.234
	Combo4	4094.6	4097.5	4096.32	4102.4	38.2	38.57	38.58	38.61	3.791	3.602	3.592	3.61
	Average	4080.6	4096	4097	4099	38.02	38.38	38.39	38.418	2.061	2.03	1.993	1.873

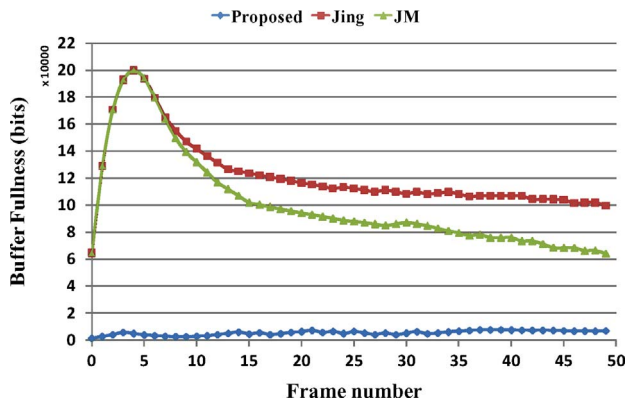


Fig. 4. Buffer fullness for *Foreman* sequence at 1024 kbps.

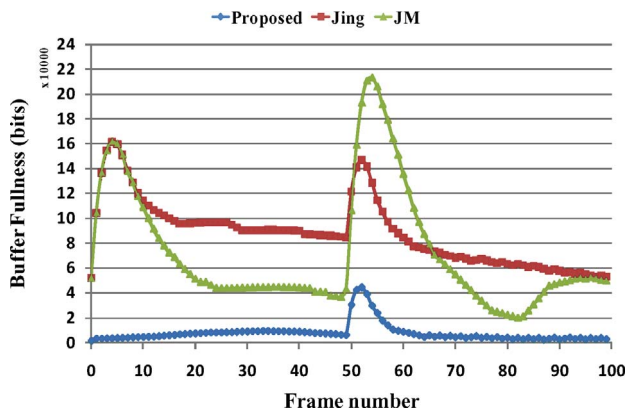


Fig. 5. Buffer fullness for *Combo2* sequence at 768 kbps.

With scene-change aware model, T+SC improves the performance not only for cascading sequences, but also for general sequences. This is due to that, in our approach, the first frame of a sequence is regarded as a scene-change frame and thus, the scene-change aware model is applied to determine its QP. Using JM’s method for QP determination of the first frame, Jing’s and JM’s methods suffer from inadequate initial QPs that may degrade the performance of beginning frames and affect overall performance. According to Table I, the T+SC method can increase average PSNR by up to 2.2 dB and reducing PSNR deviation by up to 88%, in comparison to Jing’s method and JM. As for bit-rate control, T+SC also outperforms others slightly. The results demonstrate that the proposed T+SC method can determine better QPs for both I_G and I_{SC} frames.

Figs. 4 and 5 show buffer occupancy versus frames for sequences: *Foreman* and *Combo2*, respectively. For both sequences, the proposed algorithm shows superior performance by achieving stable buffer fullness at a very low level. The reason is that, with our approach, the bit-rate generated from each frame is closely equivalent to the target bit-rate, i.e., close to instantaneous channel bit rate. Hence, the buffer fullness is kept at a stable and low level. Even though JM and Jing’s methods have average bitrates similar to the average bit-rate of the proposed method, they produced too much bits for initial few frames and those frames right after scene change (the 50th frame of *Combo2* sequence). As a consequence, the fullness of their buffers is kept at a very high level at these frames and takes long time (across many frames) to consume these bits. With stable and low buffer fullness, the proposed scheme can adopt small buffer in real-time transmission without causing buffer overflow.

IV. CONCLUSION

This letter presented a rate control algorithm for H.264/AVC video coding. We proposed a Taylor-series-based rate-QS model which was able to correctly predict the bit-rate of an intra-frame for any given QP such that the best QP subject to the target bit-rate for this frame could be determined. We also proposed a scene-change aware rate-QS model which determined QP for a scene-change frame by using the information of this frame only. The experimental results showed that our approach achieved a better performance than competed methods in regard to average PSNR, PSNR deviation, buffer fullness, and bit rate control.

REFERENCES

- [1] G. J. Sullivan, H. Yu, S. Sekiguchi, H. Sun, T. Wedi, S. Wittmann, Y. Lee, A. Segall, and T. Suzuki, "New standardized extension of MPEG-4AVC/H.264 for professional-quality video applications," in *Proc. IEEE ICIP*, vol. 1, Sep. 2007, pp. 13–16.
- [2] C.-M. Fu, W.-L. Hwang, and C.-L. Huang, "Efficient post-compression error-resilient 3-D-scalable video transmission for packet erasure channels," in *Proc. IEEE ICASSP*, Mar. 2005, pp. 305–308.
- [3] C. W. Hsiao and W. J. Tsai, "Hybrid multiple description coding based on H.264," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 1, pp. 76–87, Jan. 2010.
- [4] Z. G. Li, W. Gao, F. Pan, S. W. Ma, K. P. Lim, G. N. Feng, X. Lin, S. Rahardja, H. Q. Lu, and Y. Lu, "Adaptive rate control for H.264," *J. Vis. Commun. Image Represent.*, vol. 17, no. 2, pp. 376–406, Apr. 2006.
- [5] JM 15.0. *H.264/AVC Ref. Software* [Online]. Available: <http://iphome.hhi.de/suehring/tml/>
- [6] H. J. Lee, T. Chiang, and Y.-Q. Zhang, "Scalable rate control for MPEG-4 video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 6, pp. 878–894, Sep. 2000.
- [7] A. N. Netravali and J. O. Limb, "Picture coding: A review," *Proc. IEEE*, vol. 68, no. 3, pp. 7–12, Mar. 1960.
- [8] N. Kamaci, Y. Altunbasak, and R. M. Mersereau, "Frame bit allocation for the H.264/AVC video coder via a cauchy-density-based rate and distortion models," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 8, pp. 994–1006, Aug. 2005.
- [9] X. Jing, L.-P. Chau, and W.-C. Siu, "Frame complexity-based rate-quantization model for H.264/AVC intraframe rate control," *IEEE Signal Process. Lett.*, vol. 15, no. 1, pp. 373–376, 2008.
- [10] *Taylor Series* [Online]. Available: [http://en.wikipedia.org/wiki/Taylor\\${}_{}\\$series](http://en.wikipedia.org/wiki/Taylor${}_{}$series)
- [11] X. Jing and L.-P. Chau, "A novel intra-rate estimation method for H.264 rate control," in *Proc. ISCAS*, 2006, pp. 5019–5022.